Extending the Palm likelihood approach to inhomogeneous Neyman-Scott processes (Theory of Biomathematics and Its Applications IX)

Shimatani, Ichiro Ken

数理解析研究所講究録
第 1853 巻 2013 年 180-184

実データに直接適用可能な空間モデルとしての点過程

統計数理研究所　島谷健一郎

Extending the Palm likelihood approach to inhomogeneous Neyman-Scott processes

The Institute of Statistical Mathematics　Ichiro Ken Shimatani

1. Introduction

Conventionally, the optimization of unknown parameters in Neyman-Scott point processes have been conducted by minimizing the square difference between the observed and the expected 2nd order moment functions. The maximum Palm likelihood approach introduced by Tanaka et al. (2008) provided a new likelihood-based methodology and exhibited better performances than the conventional minimum contrast methods. The objective of this study is to extend this approach to inhomogeneous Neyman-Scott processes. This paper first reviews the likelihood equation for inhomogeneous Poisson processes and the Palm likelihood for the homogeneous Neyman-Scott processes, then proposes the extension of this method to inhomogeneous processes.

2. Inhomogeneous Poisson process

If the presence of a tree at $\mathbf{x}$ is solely determined by function $\lambda(\mathbf{x})$ which indicates the probability of the presence of a tree in the infinitesimally small unit area centered at $\mathbf{x}$, and there is no interaction between trees, the stochastic point pattern model is called the *inhomogeneous Poisson process* with the intensity function $\lambda(\mathbf{x})$. If $\lambda(\mathbf{x}) = \mu$ (constant function), the model is called the *homogeneous Poisson process* of intensity μ.

If there are trees at $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ in a rectangular plot A, the log-likelihood is given by (Cressie 1991; p. 655):

$$\sum_{i=1}^{n} \ln(\lambda(\mathbf{x}_i)) - \int_A \lambda(\mathbf{x})d\mathbf{x}. \tag{1}$$

3. Homogeneous Neyman-Scott process

Suppose that (1) a parental population followed the homogeneous Poisson process with intensity μ; (2) each parent produced a random number of offspring according to the Poisson distribution of intensity ν; (3) offspring were dispersed from each mother tree according to the two-dimensional Gaussian distribution $\exp(-r^2/2\sigma^2)/2\pi\sigma^2$; (4) there was no interaction among offspring; and (5) parents all died. The resulting distribution of offspring is called the *Thomas process*, which is contained in a more general framework of the homogeneous Neyman-Scott process.

Given the presence of a tree at location $\mathbf{x}$, the occurrence probability of another offspring at location $\mathbf{y}$ is called the *Palm intensity function.* Let this function denote as $\lambda_{\mathbf{x}}(\mathbf{y}) = P$(there is a tree at $\mathbf{y}$| presence of a tree at $\mathbf{x}$). If the point process is stationary and isotropic, $\lambda_{\mathbf{x}}(\mathbf{y})$ depends only on $r = ||\mathbf{y} - \mathbf{x}||$, thus, fixing as $\mathbf{x} = \mathbf{O}$, we may write it as $\lambda_{\mathbf{o}}(r)$. For the Thomas process, the Palm intensity function can be explicitly written as (equation (9) of Tanaka *et al.* (2008)):

$$\lambda_{\mathbf{o}}(r) = \nu\mu + \nu\cdot\exp(-r^2/4\sigma^2)/4\pi\sigma^2. \qquad (2)$$

Let R be some positive constant that is sufficiently greater than the clustering scale Suppose that trees are at $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ in plot A. Let B be the inner region that is further than R from the edges, and let $N(B)$ the number of $\mathbf{x}_i$s that belong to B. For $\mathbf{x}_i \in B$ and $\mathbf{x}_j \in A$, the *difference process* is defined as $\Delta_{ij} = \mathbf{x}_i - \mathbf{x}_j$ $(i \neq j)$ and let $r_{ij} = ||\Delta_{ij}||$. Assuming that the distribution of the difference processes can be approximated by the inhomogeneous Poisson process with intensity function $N(B)\lambda_{\mathbf{o}}(r)$, Tanaka *et al.* (2008) introduced the *log Palm likelihood* as:

$$\ln L(\mu,\nu,\sigma) = \sum_{x_i \in B}\sum_{r_{ij} \le R} \ln(N(B)\lambda_0(r_{ij})) - \int_0^R N(B)\lambda_0(r)2\pi r dr. \qquad (3)$$

(Note: Tanaka *et al.* (2008) did not separate the inner side (B) and used all the pairs for the difference process, and added the edge correction term into the equation). If the constant term, $\ln(N(B))$ is deleted, equation (3) is written as:

$$\ln \bar{L}(\mu,\nu,\sigma) = \sum_{x_i \in B}\{\sum_{r_{ij} \le R} \ln(\lambda_0(r_{ij})) - \int_0^R \lambda_0(r)2\pi r dr\} \qquad (4)$$

For the Thomas process, substituting (2), we have the explicit form;

$$\ln \bar{L}(\mu,\nu,\sigma) = \sum_{x_i \in B}\{\sum_{r_{ij} \le R} \ln\{\nu(\mu + \frac{\exp(-r_{ij}^2/4\sigma^2)}{4\pi\sigma^2})\} - \nu(\mu\pi R^2 + 1 - \exp(-R^2/4\sigma^2))\}\cdot \qquad (5)$$

Tanaka *et al.* (2008) proposed to maximize equation (5) for finding appropriate parameter values.

4. Inhomogeneous Neyman-Scott process

After the Thomas process produced an offspring population, if the location-dependent thinning operated, the resulting point pattern is an example of the *inhomogeneous Neyman-Scott process* (Waagepetersen 2007).

For this point process, the Palm intensity function can be written as:

$$\lambda_{\mathbf{x}}(\mathbf{y}) = s(\mathbf{y}; a)(\nu\mu + \nu\cdot\exp(-\|\mathbf{y} - \mathbf{x}\|^2/4\sigma^2)/4\pi\sigma^2). \tag{6}$$

In the same way as the homogeneous Neyman-Scott process, if we assume that the difference process can be approximated by the inhomogeneous Poisson process of intensity function given by equation (6), we have the log Palm likelihood for the inhomogeneous Neyman-Scott process as:

$$\ln\tilde{L}(\mu,\nu,\sigma,a) = \sum_{\mathbf{x}_i \in B}\{\sum_{\|\mathbf{x}_j-\mathbf{x}_i\|\le R}\ln(\lambda_{\mathbf{x}_i}(\mathbf{x}_j)) - \int_{\|\mathbf{y}\|\le R}\lambda_{\mathbf{x}_i}(\mathbf{y})d\mathbf{y}\}$$

$$= \sum_{\mathbf{x}_i \in B}[\sum_{\|\mathbf{x}_j-\mathbf{x}_i\|\le R}\ln\{s(\mathbf{x}_j;a)\nu(\mu+\frac{\exp(-\|\mathbf{x}_j-\mathbf{x}_i\|^2/4\sigma^2)}{4\pi\sigma^2})\} - \int_{\|\mathbf{y}-\mathbf{x}_i\|\le R}s(\mathbf{y};a)\nu(\mu+\frac{\exp(-\|\mathbf{y}-\mathbf{x}_i\|^2/4\sigma^2)}{4\pi\sigma^2})d\mathbf{y}]. \tag{7}$$

In general, the integral in equation (7) does not have an explicit form, thus, we have to numerically compute the integral, for example, by the Riemannian sum. For this purpose, a square is more convenient than a circle. Hereafter, let $\| \mathbf{y} - \mathbf{x}_i \|_S \le R$ indicate the square of edge length $2R$ centered at $\mathbf{x}_i$. Equation (7) is rewritten as:

$$\ln L_{INS}(\mu,\nu,\sigma,\Theta) = \sum_{\mathbf{x}_i \in B}[\sum_{\|\mathbf{x}_j-\mathbf{x}_i\|_S\le R}\ln\{s(\mathbf{x}_j;a)\nu(\mu+\frac{\exp(-\|\mathbf{x}_j-\mathbf{x}_i\|^2/4\sigma^2)}{4\pi\sigma^2})\} - \int_{\|\mathbf{y}-\mathbf{x}_i\|_S\le R}s(\mathbf{y};a)\nu(\mu+\frac{\exp(-\|\mathbf{y}-\mathbf{x}_i\|^2/4\sigma^2)}{4\pi\sigma^2})d\mathbf{y}]\}. \tag{8}$$

5. Example

Fig. 1 shows an example of realized point patterns of an inhomogeneous Neyman Scott process. The environmental gradient is given by

$$f(\mathbf{x}) = \sum_{j=1}^{4} e^{-\frac{(\mathbf{x}-\mathbf{z}_j)^2}{2s^2}} / 2\pi s^2 , \qquad (9)$$

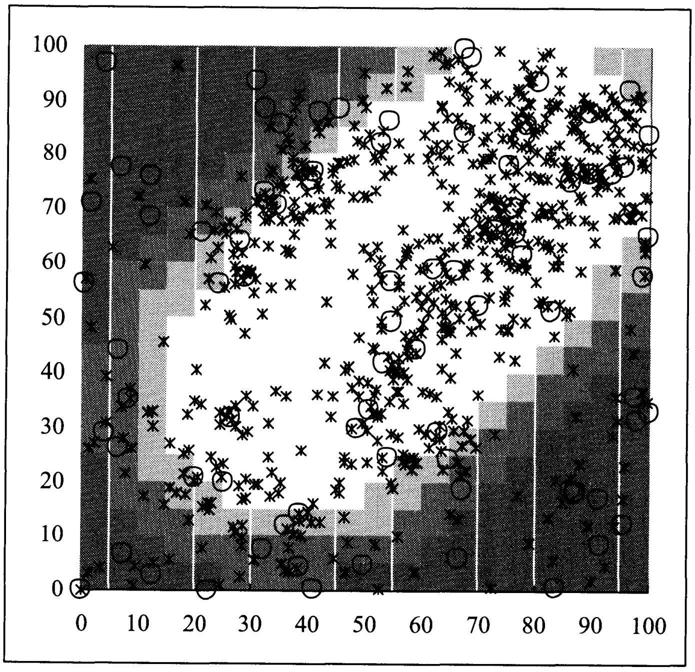
where the four (local) maxima ($\mathbf{z}_j$) are (35, 35), (50, 50), (65, 65), (80, 80), and s = 15. The survival function is given by

$$s(\mathbf{x}) = \frac{1}{1 + \exp(-8f(\mathbf{x}) + 2)} \qquad (10).$$

Parents are randomly distributed with density 0.008, and produced random numbers of offspring according to the Poisson distribution of intensity 15. Offspring were dispersed by the 2-dimensional normal distribution with variance = 15^2. For human eyes, in Fig. 1, clustering by dispersal limitation and that by high survival probabilities are mixed and hardly distinguishable.

Applying the extended maximum Palm likelihood methods, we obtained the estimates shown in Table 1. We have obtained parameter values that were close to the true ones, except the parental density for which we had 1.5 times over estimate.

Fig. 1



An example of realized point pattern by the inhomogeneous Neyman-Scott process. The circles are parents and black dots are offspring. Darker areas had lower survival probabilities.

Table 1

Parameter values obtained by the maximum Palm likelihood method. The results over 10 simulated point patterns are shown.

Parameter	True value	Mean	SD
Dispersal	5	5.609	1.588
Parental density	0.008	0.0130	0.0185
Survival probability	-8	-8.083	3.754
	2	2.056	0.525

Reference

Cressie, N. (1991) *Statistics for Spatial Data.* Wiley, New York.

Tanaka, U., Ogata, Y. & Stoyan, D. (2008) Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical Journal*, 50, 43–57.

Waagepetersen, R.P. (2007) An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 62, 252–258.